NTNU

Norwegian University of Science and Technology

The Riemannian Convex Bundle Method

Hajg Jasa

joint work with

Ronny Bergmann and Roland Herzog

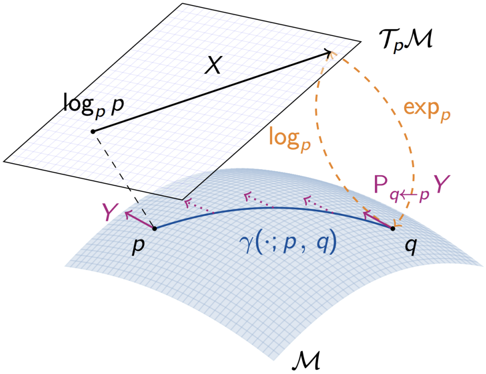
Optimization Seminar, IWR, Heidelberg

December 13th, 2023

Riemannian Geometry

Notation

- Smooth Riemannian manifold $\mathcal{M}$
- Tangent space $\mathcal{T}_p\mathcal{M}$ at the point $p \in \mathcal{M}$
- Inner product $(\cdot,\cdot)_p : \mathcal{T}_p\mathcal{M} \times \mathcal{T}_p\mathcal{M} \to \mathbb{R}$
- Exponential map $\exp_p X_p = \gamma_{pq}(1) = q$
- Logarithmic map $\log_p q = \exp_p^{-1} q = X_p$
- Parallel transport $\mathrm{P}_{q\leftarrow p} : \mathcal{T}_p\mathcal{M} \to \mathcal{T}_q\mathcal{M}$



The Problem

Consider the following minimization problem

$$\text{minimize } f(p), \quad p \in \mathcal{M},$$

where

The Problem

Consider the following minimization problem

$$\text{minimize } f(p), \quad p \in \mathcal{M},$$

where

- $\mathcal{M}$ is a Riemannian manifold with the following properties:

The Problem

Consider the following minimization problem

$$\operatorname{minimize} f(p), \quad p \in \mathcal{M},$$

where

- $\mathcal{M}$ is a Riemannian manifold with the following properties:
  - complete
  - bounded sectional curvature on $\operatorname{dom} f = \{p \in \mathcal{M} \mid f(p) < +\infty\}$, i. e. there exist $\omega, \Omega \in \mathbb{R}$ such that the sectional curvature κ satisfies $\omega \leq \kappa(p) \leq \Omega$ for all $p \in \operatorname{dom} f$.

The Problem

Consider the following minimization problem

$$\operatorname{minimize} f(p), \quad p \in \mathcal{M},$$

where

- $\mathcal{M}$ is a Riemannian manifold with the following properties:
  - complete
  - bounded sectional curvature on $\operatorname{dom} f = \{p \in \mathcal{M} \mid f(p) < +\infty\}$, i. e. there exist $\omega, \Omega \in \mathbb{R}$ such that the sectional curvature κ satisfies $\omega \leq \kappa(p) \leq \Omega$ for all $p \in \operatorname{dom} f$.
- $f\colon \mathcal{M} \to \overline{\mathbb{R}}$ is such that

The Problem

Consider the following minimization problem

$$\operatorname{minimize} f(p), \quad p \in \mathcal{M},$$

where

- $\mathcal{M}$ is a Riemannian manifold with the following properties:
  - complete
  - bounded sectional curvature on $\operatorname{dom} f = \{p \in \mathcal{M} \mid f(p) < +\infty\}$, i. e. there exist $\omega, \Omega \in \mathbb{R}$ such that the sectional curvature κ satisfies $\omega \leq \kappa(p) \leq \Omega$ for all $p \in \operatorname{dom} f$.
- $f\colon \mathcal{M} \to \overline{\mathbb{R}}$ is such that
  - $\operatorname{dom} f$ is a nonempty, bounded, strongly geodesically convex set with nonempty interior in $\mathcal{M}$
  - geodesically convex on $\operatorname{dom} f$: $f \circ \gamma\colon [0, 1] \to \overline{\mathbb{R}}$ for all geodesic arcs $\gamma \subseteq \operatorname{dom} f$ is convex in the usual sense
  - lower semi-continuous: $\liminf_{q \to p} f(q) \geq f(p)$

The Problem

Consider the following minimization problem

$$\operatorname{minimize} f(p), \quad p \in \mathcal{M},$$

where

- $\mathcal{M}$ is a Riemannian manifold with the following properties:
  - complete
  - bounded sectional curvature on $\operatorname{dom} f = \{p \in \mathcal{M} \mid f(p) < +\infty\}$, i. e. there exist $\omega, \Omega \in \mathbb{R}$ such that the sectional curvature κ satisfies $\omega \leq \kappa(p) \leq \Omega$ for all $p \in \operatorname{dom} f$.
- $f\colon \mathcal{M} \to \overline{\mathbb{R}}$ is such that
  - $\operatorname{dom} f$ is a nonempty, bounded, strongly geodesically convex set with nonempty interior in $\mathcal{M}$
  - geodesically convex on $\operatorname{dom} f$: $f \circ \gamma\colon [0,1] \to \overline{\mathbb{R}}$ for all geodesic arcs $\gamma \subseteq \operatorname{dom} f$ is convex in the usual sense
  - lower semi-continuous: $\liminf_{q \to p} f(q) \geq f(p)$

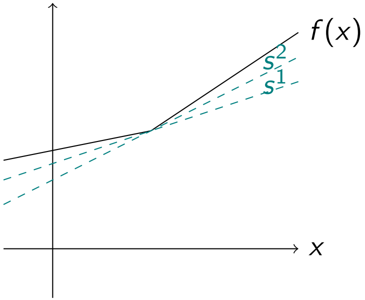
Goal. Solve this non-smooth optimization problem with a convex bundle method.

The Convex Subdifferential(s)

For a convex function, the subdifferential is defined as

$$\partial f(x) = \left\{ s \in \mathbb{R}^n \;\middle|\; f(y) \geq f(x) + (s)^T (y - x) \text{ for all } y \in \mathbb{R}^n \right\}$$

and it is a non-empty, closed and convex subset.



The Convex Subdifferential(s)

For a convex function, the subdifferential is defined as

$$\partial f(x) = \left\{ s \in \mathbb{R}^n \,\middle|\, f(y) \geq f(x) + (s)^T (y - x) \text{ for all } y \in \mathbb{R}^n \right\}$$

and it is a non-empty, closed and convex subset. Let $\varepsilon > 0$. The $\varepsilon-$subdifferential is

$$\partial_\varepsilon f(x) = \left\{ s \in \mathbb{R}^n \,\middle|\, f(y) \geq f(x) + (s)^T (y - x) - \varepsilon \text{ for all } y \in \mathbb{R}^n \right\}$$

and

$$\partial f(x) \subseteq \partial_\varepsilon f(x)$$

The Convex Subdifferential(s)

For a geodesically convex function, the subdifferential is defined as

$$\partial f(p) = \bigl\{ X_p \in \mathcal{T}_p\mathcal{M} \bigm| f(q) \geq f(p) + \bigl(X_p, \log_p q\bigr) \text{ for all } q \in \operatorname{dom} f \bigr\}$$

and it is a non-empty, closed and convex subset. Let $\varepsilon > 0$. The $\varepsilon-$subdifferential is

$$\partial_\varepsilon f(p) = \bigl\{ X_p \in \mathcal{T}_p\mathcal{M} \bigm| f(q) \geq f(p) + \bigl(X_p, \log_p q\bigr) - \varepsilon \text{ for all } q \in \operatorname{dom} f \bigr\}$$

and

$$\partial f(p) \subseteq \partial_\varepsilon f(p)$$

The Bundle Method

Bundle methods are about descent as well as stability.

The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of a general bundle algorithm (Bonnans et al. 2006):

The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of a general bundle algorithm (Bonnans et al. 2006):

- keeps memory of the last “best” points

The Bundle Method

NTNU

Bundle methods are about descent as well as stability.

Core characteristics of a general bundle algorithm (Bonnans et al. 2006):

- keeps memory of the last “best” points
- solves a stabilization subproblem

The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of a general bundle algorithm (Bonnans et al. 2006):

- keeps memory of the last “best” points
- solves a stabilization subproblem
    - we employ a dual approach with the $\varepsilon-$subdifferential as in Geiger and Kanzow 2002

The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of a general bundle algorithm (Bonnans et al. 2006):

- keeps memory of the last “best” points
- solves a stabilization subproblem
  - we employ a dual approach with the $\varepsilon-$subdifferential as in Geiger and Kanzow 2002
- generates sequences of *candidate* points and *stability centers*

Approximating the $\varepsilon-$subdifferential

Given $x^1, \ldots, x^k \in \mathbb{R}^n$, and $s^j \in \partial f(x^j)$, then

Approximating the $\varepsilon-$subdifferential

Given $x^1, \ldots, x^k \in \mathbb{R}^n$, and $s^j \in \partial f(x^j)$, then $s^j \in \partial_{e_j^k} f(x^k)$

Approximating the $\varepsilon-$subdifferential

Given $x^1, \ldots, x^k \in \mathbb{R}^n$, and $s^j \in \partial f(x^j)$, then $s^j \in \partial_{e_j^k} f(x^k)$

with $e_j^k := f(x^k) - f(x^j) - (s^j)^T(x^k - x^j)$ for all $j \in \{1, \ldots, k\}$

Approximating the $\varepsilon-$subdifferential

Given $x^1, \dots, x^k \in \mathbb{R}^n$, and $s^j \in \partial f(x^j)$, then $s^j \in \partial_{e_j^k} f(x^k)$

with $e_j^k := f(x^k) - f(x^j) - (s^j)^T(x^k - x^j)$ for all $j \in \{1, \dots, k\}$

We can characterize an inner approximation of $\partial_\varepsilon f(p)$ as:

$$G_\varepsilon^k := \left\{ \sum_{j=1}^{k} \lambda_j s^j \;\middle|\; \sum_{j=1}^{k} \lambda_j \, e_j^k \leq \varepsilon, \; \sum_{j=1}^{k} \lambda_j = 1, \; \lambda_j \geq 0 \text{ for all } j = 1, \dots, k \right\}$$

with $G_\varepsilon^k \subseteq \partial_\varepsilon f(x^k)$.

Approximating the $\varepsilon-$subdifferential

Given $x^1, \dots, x^k \in \mathbb{R}^n$, and $s^j \in \partial f(x^j)$, then $s^j \in \partial_{e_j^k} f(x^k)$

$$\text{with} \quad e_j^k := f(x^k) - f(x^j) - (s^j)^T(x^k - x^j) \quad \text{for all} \quad j \in \{1, \dots, k\}$$

We can characterize an inner approximation of $\partial_\varepsilon f(p)$ as:

$$G_\varepsilon^k := \left\{ \sum_{j=1}^k \lambda_j s^j \,\middle|\, \sum_{j=1}^k \lambda_j \, e_j^k \leq \varepsilon, \; \sum_{j=1}^k \lambda_j = 1, \; \lambda_j \geq 0 \text{ for all } j = 1, \dots, k \right\}$$

with $G_\varepsilon^k \subseteq \partial_\varepsilon f(x^k)$.

- Main challenge on manifolds: given $p^1, \dots, p^k \in \mathcal{M}$, and $X_{p^j} \in \partial f(p^j)$, then

$$\mathrm{P}_{p^k \leftarrow p^j} X_{p^j} \in \partial_c f(p^k) \quad \text{for some} \quad c > 0?$$

Curvature Correction

Let $\delta := \operatorname{diam}(\operatorname{dom} f) < +\infty$. Define

$$\zeta_{1,\omega}(\delta) := \begin{cases} 1 & \text{if } \omega \geq 0, \\ \sqrt{-\omega}\,\delta \coth(\sqrt{-\omega}\,\delta) & \text{if } \omega < 0, \end{cases}$$

$$\zeta_{2,\Omega}(\delta) := \begin{cases} 1 & \text{if } \Omega \leq 0, \\ \sqrt{\Omega}\,\delta \cot(\sqrt{\Omega}\,\delta) & \text{if } \Omega > 0, \end{cases}$$

Curvature Correction

Let $\delta := \operatorname{diam}(\operatorname{dom} f) < +\infty$. Define

$$\zeta_{1,\omega}(\delta) := \begin{cases} 1 & \text{if } \omega \geq 0, \\ \sqrt{-\omega}\,\delta \coth(\sqrt{-\omega}\,\delta) & \text{if } \omega < 0, \end{cases}$$

$$\zeta_{2,\Omega}(\delta) := \begin{cases} 1 & \text{if } \Omega \leq 0, \\ \sqrt{\Omega}\,\delta \cot(\sqrt{\Omega}\,\delta) & \text{if } \Omega > 0, \end{cases}$$

and $\varrho := \max\{\zeta_{1,\omega}(\delta) - 1,\ 1 - \zeta_{2,\Omega}(\delta)\}$, as in Alimisis et al. 2021, Appendix C.

Curvature Correction

Let $\delta := \operatorname{diam}(\operatorname{dom} f) < +\infty$. Define

$$\zeta_{1,\omega}(\delta) := \begin{cases} 1 & \text{if } \omega \geq 0, \\ \sqrt{-\omega}\,\delta \coth(\sqrt{-\omega}\,\delta) & \text{if } \omega < 0, \end{cases}$$

$$\zeta_{2,\Omega}(\delta) := \begin{cases} 1 & \text{if } \Omega \leq 0, \\ \sqrt{\Omega}\,\delta \cot(\sqrt{\Omega}\,\delta) & \text{if } \Omega > 0, \end{cases}$$

and $\varrho := \max\bigl\{\zeta_{1,\omega}(\delta) - 1,\ 1 - \zeta_{2,\Omega}(\delta)\bigr\}$, as in Alimisis et al. 2021, Appendix C. By defining

$$c_j^k := f(p^k) - f(p^j) - \Bigl(X_{p^j}, \log_{p^j} p^k\Bigr) + \varrho \|\log_{p^j} p^k\| \|X_{p^j}\|,$$

with $X_{p_j} \in \partial f(p^j)$,

Curvature Correction

Let $\delta := \operatorname{diam}(\operatorname{dom} f) < +\infty$. Define

$$\zeta_{1,\omega}(\delta) := \begin{cases} 1 & \text{if } \omega \geq 0, \\ \sqrt{-\omega}\,\delta \coth(\sqrt{-\omega}\,\delta) & \text{if } \omega < 0, \end{cases}$$

$$\zeta_{2,\Omega}(\delta) := \begin{cases} 1 & \text{if } \Omega \leq 0, \\ \sqrt{\Omega}\,\delta \cot(\sqrt{\Omega}\,\delta) & \text{if } \Omega > 0, \end{cases}$$

and $\varrho := \max\bigl\{\zeta_{1,\omega}(\delta) - 1,\ 1 - \zeta_{2,\Omega}(\delta)\bigr\}$, as in Alimisis et al. 2021, Appendix C. By defining

$$c_j^k := f(p^k) - f(p^j) - \Bigl(X_{p^j}, \log_{p^j} p^k\Bigr) + \varrho \|\log_{p^j} p^k\| \|X_{p^j}\|,$$

with $X_{p_j} \in \partial f(p^j)$, we get

$$G_\varepsilon^k := \left\{ \sum_{j=1}^k \lambda_j \mathrm{P}_{p^k \leftarrow p^j} X_{p^j} \;\middle|\; \sum_{j=1}^k \lambda_j\, c_j^k \leq \varepsilon,\ \sum_{j=1}^k \lambda_j = 1,\ \lambda_j \geq 0 \text{ for all } j = 1, \ldots, k \right\}$$

with $G_\varepsilon^k \subseteq \partial_\varepsilon f(p^k)$, and $\mathrm{P}_{p^k \leftarrow p^j} X_{p^j} \in \partial_{c_j^k} f(p^k)$.

The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \ldots, k\} = J_k$ and let λ_j be convex coefficients.

The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \dots, k\} = J_k$ and let λ_j be convex coefficients. In the Euclidean case:

$$s^j \in \partial f(x^j) \implies \sum_{j=1}^{k} \lambda_j s^j \in \partial_\varepsilon f(x^k)$$

The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \dots, k\} = J_k$ and let λ_j be convex coefficients. In the Euclidean case:

$$s^j \in \partial f(x^j) \implies \sum_{j=1}^{k} \lambda_j s^j \in \partial_\varepsilon f(x^k) \quad \text{iff} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j e_j^k$$

The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \dots, k\} = J_k$ and let λ_j be convex coefficients. In the Euclidean case:

$$s^j \in \partial f(x^j) \implies \sum_{j=1}^{k} \lambda_j s^j \in \partial_\varepsilon f(x^k) \quad \text{iff} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j e_j^k$$

The *search direction* is then the projection of the zero vector onto G_ε^k. Due to the characterization of this set, this translates to

$$d^k := -\sum_{j=1}^{k} \lambda_j s^j$$

The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \ldots, k\} = J_k$ and let λ_j be convex coefficients. In the Euclidean case:

$$s^j \in \partial f(x^j) \implies \sum_{j=1}^{k} \lambda_j s^j \in \partial_\varepsilon f(x^k) \quad \text{iff} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j e_j^k$$

The *search direction* is then the projection of the zero vector onto G_ε^k. Due to the characterization of this set, this translates to

$$d^k := -\sum_{j=1}^{k} \lambda_j s^j$$

where the coefficients λ_j are solution to

The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \dots, k\} = J_k$ and let λ_j be convex coefficients. In the Euclidean case:

$$s^j \in \partial f(x^j) \implies \sum_{j=1}^{k} \lambda_j s^j \in \partial_\varepsilon f(x^k) \quad \text{iff} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j e_j^k$$

The *search direction* is then the projection of the zero vector onto G_ε^k. Due to the characterization of this set, this translates to

$$d^k := -\sum_{j=1}^{k} \lambda_j s^j$$

where the coefficients λ_j are solution to

$$\begin{aligned} \underset{\lambda \in \mathbb{R}^{|J_k|}}{\arg\min} \quad & \frac{1}{2} \left\| \sum_{j \in J_k} \lambda_j s^j \right\|^2 + \sum_{j \in J_k} \lambda_j e_j^k \\ \text{s. t.} \quad & \sum_{j \in J_k} \lambda_j = 1, \quad \lambda_j \geq 0 \qquad \text{for all } j \in J_k \end{aligned}$$

The Riemannian Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \dots, k\} = J_k$ and let λ_j be convex coefficients. In the Riemannian case:

$$s^j \in \partial f(x^j) \implies \sum_{j=1}^{k} \lambda_j s^j \in \partial_\varepsilon f(x^k) \quad \text{iff} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j e_j^k$$

The *search direction* is then the projection of the zero vector onto G_ε^k. Due to the characterization of this set, this translates to

$$d^k := -\sum_{j=1}^{k} \lambda_j s^j$$

where the coefficients λ_j are solution to

$$\begin{aligned} \operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{|J_k|}} \quad & \frac{1}{2} \left\| \sum_{j \in J_k} \lambda_j s^j \right\|^2 + \sum_{j \in J_k} \lambda_j e_j^k \\ \text{s. t.} \quad & \sum_{j \in J_k} \lambda_j = 1, \quad \lambda_j \geq 0 \qquad \text{for all } j \in J_k \end{aligned}$$

The Riemannian Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \dots, k\} = J_k$ and let λ_j be convex coefficients. In the Riemannian case:

$$X_{q^j} \in \partial f(q^j) \implies \sum_{j=1}^{k} \lambda_j \mathrm{P}_{p^k \leftarrow q^j} X_{q^j} \in \partial_\varepsilon f(p^k) \quad \text{iff} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j c_j^k$$

The *search direction* is then the projection of the zero vector onto G_ε^k. Due to the characterization of this set, this translates to

$$d^k := -\sum_{j=1}^{k} \lambda_j s^j$$

where the coefficients λ_j are solution to

$$\begin{aligned} &\operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{|J_k|}} && \frac{1}{2} \Big\| \sum_{j \in J_k} \lambda_j s^j \Big\|^2 + \sum_{j \in J_k} \lambda_j e_j^k \\ &\text{s. t.} && \sum_{j \in J_k} \lambda_j = 1, \quad \lambda_j \geq 0 \qquad \text{for all } j \in J_k \end{aligned}$$

The Riemannian Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \ldots, k\} = J_k$ and let λ_j be convex coefficients. In the Riemannian case:

$$X_{q^j} \in \partial f(q^j) \implies \sum_{j=1}^{k} \lambda_j \mathrm{P}_{p^k \leftarrow q^j} X_{q^j} \in \partial_\varepsilon f(p^k) \quad \text{iff} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j c_j^k$$

The *search direction* is then the projection of the zero vector onto G_ε^k. Due to the characterization of this set, this translates to

$$d^k := -\sum_{j=1}^{k} \lambda_j \mathrm{P}_{p^k \leftarrow q^j} X_{q^j}$$

where the coefficients λ_j are solution to

$$\begin{aligned} \operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{|J_k|}} \quad & \frac{1}{2} \Bigg\| \sum_{j \in J_k} \lambda_j s^j \Bigg\|^2 + \sum_{j \in J_k} \lambda_j e_j^k \\ \text{s. t.} \quad & \sum_{j \in J_k} \lambda_j = 1, \quad \lambda_j \geq 0 \qquad \text{for all } j \in J_k \end{aligned}$$

The Riemannian Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \dots, k\} = J_k$ and let λ_j be convex coefficients. In the Riemannian case:

$$X_{q^j} \in \partial f(q^j) \implies \sum_{j=1}^{k} \lambda_j \mathrm{P}_{p^k \leftarrow q^j} X_{q^j} \in \partial_\varepsilon f(p^k) \quad \text{iff} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j c_j^k$$

The *search direction* is then the projection of the zero vector onto G_ε^k. Due to the characterization of this set, this translates to

$$d^k := -\sum_{j=1}^{k} \lambda_j \mathrm{P}_{p^k \leftarrow q^j} X_{q^j}$$

where the coefficients λ_j are solution to

$$\begin{aligned} \operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{|J_k|}} \quad & \frac{1}{2} \Big\lVert \sum_{j \in J_k} \lambda_j \mathrm{P}_{p^k \leftarrow q^j} X_{q^j} \Big\rVert^2 + \sum_{j \in J_k} \lambda_j c_j^k \\ \text{s. t.} \quad & \sum_{j \in J_k} \lambda_j = 1, \quad \lambda_j \geq 0 \qquad \text{for all } j \in J_k \end{aligned}$$

The Euclidean Convex Bundle Method

Data: $x^1 = y^1 \in \mathbb{R}^n$, $g^0 = s^1 \in \partial f(x^1)$, $m \in (0,1)$, $\varepsilon_0 = e_1^1 = 0$, $k = 1$, $J_k = \{1\}$

while *not converged* **do**

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the subproblem and set

$$g^k := \sum_{j \in J_k} \lambda_j^k s^j, \quad \varepsilon_k := \sum_{j \in J_k} \lambda_j^k e_j^k, \quad d^k := -g^k, \quad \xi_k := -\|g^k\|^2 - \varepsilon_k$$

Set $y^{k+1} := x^k + d^k$

If $f(y^{k+1}) \leq f(x^k) + m\xi_k$, then set $x^{k+1} := y^{k+1}$, else set $x^{k+1} := x^k$

Get $s^{k+1} \in \partial f(y^{k+1})$, update the index set to J_{k+1}, and set

$$e_j^{k+1} := f(x^{k+1}) - f(y^j) - (s^j)^T (x^{k+1} - y^j) \text{ for all } j \in J_{k+1}$$

end

Result: x^{k_*}, for some $k_* \in \mathbb{N}$.

The Riemannian Convex Bundle Method

Data: $p^1 = q^1 \in \operatorname{int}(\operatorname{dom} f)$, $g^0 = X_1 \in \partial f(p^1)$, $m \in (0,1)$, $\varepsilon_0 = c_1^1 = 0$, $k = 1$, $J_k = \{1\}$, backtracking parameter $\beta > 0$

while *not converged* **do**

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the subproblem and set

$$g^k := \sum_{j \in J_k} \lambda_j^k s^j, \quad \varepsilon_k := \sum_{j \in J_k} \lambda_j^k e_j^k, \quad d^k := -g^k, \quad \xi_k := -\|g^k\|^2 - \varepsilon_k$$

Set $y^{k+1} := x^k + d^k$

If $f(y^{k+1}) \leq f(x^k) + m\xi_k$, then set $x^{k+1} := y^{k+1}$, else set $x^{k+1} := x^k$

Get $s^{k+1} \in \partial f(y^{k+1})$, update the index set to J_{k+1}, and set

$$e_j^{k+1} := f(x^{k+1}) - f(y^j) - (s^j)^T (x^{k+1} - y^j) \text{ for all } j \in J_{k+1}$$

end

Result: x^{k_*}, for some $k_* \in \mathbb{N}$.

The Riemannian Convex Bundle Method

Data: $p^1 = q^1 \in \mathrm{int}(\mathrm{dom}\, f)$, $g^0 = X_1 \in \partial f(p^1)$, $m \in (0,1)$, $\varepsilon_0 = c_1^1 = 0$,
$k = 1$, $J_k = \{1\}$, backtracking parameter $\beta > 0$

while *not converged* **do**

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the subproblem and set

$$g^k := \sum_{j \in J_k} \lambda_j^k \mathrm{P}_{p^k \leftarrow q^j} X_{q^j}, \quad \varepsilon_k := \sum_{j \in J_k} \lambda_j^k c_j^k, \quad d^k := -g^k, \quad \xi_k := -\|g^k\|^2 - \varepsilon_k$$

Set $y^{k+1} := x^k + d^k$

If $f(y^{k+1}) \leq f(x^k) + m\xi_k$, then set $x^{k+1} := y^{k+1}$, else set $x^{k+1} := x^k$

Get $s^{k+1} \in \partial f(y^{k+1})$, update the index set to J_{k+1}, and set

$$e_j^{k+1} := f(x^{k+1}) - f(y^j) - (s^j)^T (x^{k+1} - y^j) \text{ for all } j \in J_{k+1}$$

end

Result: x^{k_*}, for some $k_* \in \mathbb{N}$.

The Riemannian Convex Bundle Method

Data: $p^1 = q^1 \in \operatorname{int}(\operatorname{dom} f)$, $g^0 = X_1 \in \partial f(p^1)$, $m \in (0,1)$, $\varepsilon_0 = c_1^1 = 0$,
$k = 1$, $J_k = \{1\}$, backtracking parameter $\beta > 0$

while *not converged* **do**

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the subproblem and set

$$g^k := \sum_{j \in J_k} \lambda_j^k \mathrm{P}_{p^k \leftarrow q^j} X_{q^j}, \quad \varepsilon_k := \sum_{j \in J_k} \lambda_j^k c_j^k, \quad d^k := -g^k, \quad \xi_k := -\|g^k\|^2 - \varepsilon_k$$

Set $t_k := 1$. While $q^{k+1} := \exp_{p^k}(t_k d^k) \notin \operatorname{int}(\operatorname{dom} f)$ and $q^{k+1} \neq p^k$
backtrack $t_k = \beta t_k$

If $f(y^{k+1}) \leq f(x^k) + m\xi_k$, then set $x^{k+1} := y^{k+1}$, else set $x^{k+1} := x^k$

Get $s^{k+1} \in \partial f(y^{k+1})$, update the index set to J_{k+1}, and set

$$e_j^{k+1} := f(x^{k+1}) - f(y^j) - (s^j)^T (x^{k+1} - y^j) \text{ for all } j \in J_{k+1}$$

end

Result: x^{k_*}, for some $k_* \in \mathbb{N}$.

The Riemannian Convex Bundle Method

Data: $p^1 = q^1 \in \operatorname{int}(\operatorname{dom} f)$, $g^0 = X_1 \in \partial f(p^1)$, $m \in (0,1)$, $\varepsilon_0 = c_1^1 = 0$, $k = 1$, $J_k = \{1\}$, backtracking parameter $\beta > 0$

while *not converged* **do**

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the subproblem and set

$$g^k := \sum_{j \in J_k} \lambda_j^k \mathrm{P}_{p^k \leftarrow q^j} X_{q^j}, \quad \varepsilon_k := \sum_{j \in J_k} \lambda_j^k c_j^k, \quad d^k := -g^k, \quad \xi_k := -\|g^k\|^2 - \varepsilon_k$$

Set $t_k := 1$. While $q^{k+1} := \exp_{p^k}(t_k d^k) \notin \operatorname{int}(\operatorname{dom} f)$ and $q^{k+1} \neq p^k$ backtrack $t_k = \beta t_k$

If $f(q^{k+1}) \leq f(p^k) + m\xi_k$, then set $p^{k+1} := q^{k+1}$, else set $p^{k+1} := p^k$

Get $s^{k+1} \in \partial f(y^{k+1})$, update the index set to J_{k+1}, and set

$$e_j^{k+1} := f(x^{k+1}) - f(y^j) - (s^j)^T (x^{k+1} - y^j) \text{ for all } j \in J_{k+1}$$

end

Result: x^{k_*}, for some $k_* \in \mathbb{N}$.

The Riemannian Convex Bundle Method

Data: $p^1 = q^1 \in \operatorname{int}(\operatorname{dom} f)$, $g^0 = X_1 \in \partial f(p^1)$, $m \in (0,1)$, $\varepsilon_0 = c_1^1 = 0$, $k = 1$, $J_k = \{1\}$, backtracking parameter $\beta > 0$

while *not converged* **do**

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the subproblem and set

$$g^k := \sum_{j \in J_k} \lambda_j^k \mathrm{P}_{p^k \leftarrow q^j} X_{q^j}, \quad \varepsilon_k := \sum_{j \in J_k} \lambda_j^k c_j^k, \quad d^k := -g^k, \quad \xi_k := -\|g^k\|^2 - \varepsilon_k$$

Set $t_k := 1$. While $q^{k+1} := \exp_{p^k}(t_k d^k) \notin \operatorname{int}(\operatorname{dom} f)$ and $q^{k+1} \neq p^k$ backtrack $t_k = \beta t_k$

If $f(q^{k+1}) \leq f(p^k) + m\xi_k$, then set $p^{k+1} := q^{k+1}$, else set $p^{k+1} := p^k$

Get $X_{q^{k+1}} \in \partial f(q^{k+1})$, update the index set to J_{k+1}, and set

$$c_j^{k+1} := f(p^{k+1}) - f(q^j) - \left(X_{q^j}, \log_{q^j} p^{k+1} \right) \text{ for all } j \in J_{k+1}$$

end

Result: x^{k_*}, for some $k_* \in \mathbb{N}$.

The Riemannian Convex Bundle Method

Data: $p^1 = q^1 \in \operatorname{int}(\operatorname{dom} f)$, $g^0 = X_1 \in \partial f(p^1)$, $m \in (0,1)$, $\varepsilon_0 = c_1^1 = 0$,
$k = 1$, $J_k = \{1\}$, backtracking parameter $\beta > 0$

while *not converged* **do**

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the subproblem and set

$$g^k := \sum_{j \in J_k} \lambda_j^k \mathrm{P}_{p^k \leftarrow q^j} X_{q^j}, \quad \varepsilon_k := \sum_{j \in J_k} \lambda_j^k c_j^k, \quad d^k := -g^k, \quad \xi_k := -\|g^k\|^2 - \varepsilon_k$$

Set $t_k := 1$. While $q^{k+1} := \exp_{p^k}(t_k d^k) \notin \operatorname{int}(\operatorname{dom} f)$ and $q^{k+1} \neq p^k$ backtrack $t_k = \beta t_k$

If $f(q^{k+1}) \leq f(p^k) + m\xi_k$, then set $p^{k+1} := q^{k+1}$, else set $p^{k+1} := p^k$

Get $X_{q^{k+1}} \in \partial f(q^{k+1})$, update the index set to J_{k+1}, and set

$$c_j^{k+1} := f(p^{k+1}) - f(q^j) - \left(X_{q^j}, \log_{q^j} p^{k+1} \right) + \varrho \|\log_{q^j} p^{k+1}\| \|X_{q^j}\| \text{ for all } j \in J_{k+1}$$

end

Result: x^{k_*}, for some $k_* \in \mathbb{N}$.

The Riemannian Convex Bundle Method

Data: $p^1 = q^1 \in \operatorname{int}(\operatorname{dom} f)$, $g^0 = X_1 \in \partial f(p^1)$, $m \in (0,1)$, $\varepsilon_0 = c_1^1 = 0$, $k = 1$, $J_k = \{1\}$, backtracking parameter $\beta > 0$

while *not converged* **do**

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the subproblem and set

$$g^k := \sum_{j \in J_k} \lambda_j^k \mathrm{P}_{p^k \leftarrow q^j} X_{q^j}, \quad \varepsilon_k := \sum_{j \in J_k} \lambda_j^k c_j^k, \quad d^k := -g^k, \quad \xi_k := -\|g^k\|^2 - \varepsilon_k$$

Set $t_k := 1$. While $q^{k+1} := \exp_{p^k}(t_k d^k) \notin \operatorname{int}(\operatorname{dom} f)$ and $q^{k+1} \neq p^k$ backtrack $t_k = \beta t_k$

If $f(q^{k+1}) \leq f(p^k) + m\xi_k$, then set $p^{k+1} := q^{k+1}$, else set $p^{k+1} := p^k$

Get $X_{q^{k+1}} \in \partial f(q^{k+1})$, update the index set to J_{k+1}, and set

$$c_j^{k+1} := f(p^{k+1}) - f(q^j) - \left(X_{q^j}, \log_{q^j} p^{k+1} \right) + \varrho \|\log_{q^j} p^{k+1}\| \|X_{q^j}\| \text{ for all } j \in J_{k+1}$$

end

Result: p^{k_*}, for some $k_* \in \mathbb{N}$.

Convergence

- In the Euclidean case we have Geiger and Kanzow 2002, Theorem 6.80.

Theorem

Let the solution set $S = \{x^* \in \mathbb{R}^n \mid f(x^*) = \inf_{x \in \mathbb{R}^n} f(x)\}$ *of the minimization problem be nonempty. Then every sequence* $\{x^k\}$ *generated by the bundle method algorithm converges to a minimizer of* f.

Convergence

- In the Euclidean case we have Geiger and Kanzow 2002, Theorem 6.80.

Theorem
Let the solution set $S = \{x^* \in \mathbb{R}^n \mid f(x^*) = \inf_{x \in \mathbb{R}^n} f(x)\}$ *of the minimization problem be nonempty. Then every sequence* $\{x^k\}$ *generated by the bundle method algorithm converges to a minimizer of* f*.*

- In the Riemannian case, we have obtained convergence in case the number of serious steps generated by the algorithm is infinite.

Convergence

- In the Euclidean case we have Geiger and Kanzow 2002, Theorem 6.80.

Theorem

Let the solution set $S = \{x^* \in \mathbb{R}^n \mid f(x^*) = \inf_{x \in \mathbb{R}^n} f(x)\}$ *of the minimization problem be nonempty. Then every sequence* $\{x^k\}$ *generated by the bundle method algorithm converges to a minimizer of* f.

- In the Riemannian case, we have obtained convergence in case the number of serious steps generated by the algorithm is infinite.
- The case where only a finite number of serious steps is generated is still *open*. The main problem is studying the convergence of the curvature-dependent term $\varrho \|\log_{q^j} p^{k+1}\| \|X_{q^j}\|$.

Convergence

In Geiger and Kanzow 2002, Theorem 6.80, they are able to show that

$$\sum_{j=k_*}^{+\infty} \left(\|g^j\|^2 + \varepsilon_j\right)^2 < +\infty$$

where $k_* \in \mathbb{N}$ is the index that corresponds to the last serious iterate, namely

$$p^k = p^{k_*} \quad \text{for all } k \geq k_*.$$

This is possible because from the definition of e_j^k one gets a bound

$$\left(s^k\,,\, g^{k-1}\right) < m\left(\|g^{k-1}\|^2 + \varepsilon_{k-1}\right) - e_k^k,$$

whereas in the Riemannian case one has

$$\left(\mathrm{P}_{p^k \leftarrow q^k} X_{q^k}\,,\, t_{k-1} g^{k-1}\right) < m\left(\|g^{k-1}\|^2 + \varepsilon_{k-1}\right) - c_k^k + \varrho \,\|t_{k-1} g^{k-1}\|\,\|X_{q^k}\|.$$

Implementation

The algorithm is implemented[1] in Julia using `Manopt.jl` (Bergmann 2022) which uses manifolds from `Manifolds.jl` (Axen et al. 2021). A solver call just looks like

```
p* = convex_bundle_method(M, f, ∂f, p0)
```

where

- `M` is a Riemannian manifold
- `f` is the objective function
- `∂f` is a subgradient of the objective function
- `p0` is an initial point on the manifold
- the parameter for the descent test is set at a default $m = 0.0125$

The default stopping criterion for the algorithm is set to

$$-\xi_k \leq 10^{-8}.$$

[1] not yet in a release version

Numerical Example: Signal Denoising on $\mathcal{H}^2$

Let $\mathcal{H}^2$ be the two-dimensional hyperbolic space and let $n = 496$. By projecting a square wave onto $\mathcal{H}^2$, we manufacture an artificial signal which can be interpreted as a point $q \in (\mathcal{H}^2)^n$.

Numerical Example: Signal Denoising on $\mathcal{H}^2$

Let $\mathcal{H}^2$ be the two-dimensional hyperbolic space and let $n = 496$. By projecting a square wave onto $\mathcal{H}^2$, we manufacture an artificial signal which can be interpreted as a point $q \in (\mathcal{H}^2)^n$. We then generate a noisy signal $\bar{q} \in \mathcal{M}$ by adding Gaussian noise to the original signal, i. e., we set $\bar{q}^i = \exp_{q^i} X_i$ for $i = 1, \ldots, n$, where $X \in \mathcal{T}_{q^i}\mathcal{M}$ has a standard deviation of $\sigma = 0.1$.

Numerical Example: Signal Denoising on $\mathcal{H}^2$

Let $\mathcal{H}^2$ be the two-dimensional hyperbolic space and let $n = 496$. By projecting a square wave onto $\mathcal{H}^2$, we manufacture an artificial signal which can be interpreted as a point $q \in (\mathcal{H}^2)^n$. We then generate a noisy signal $\bar{q} \in \mathcal{M}$ by adding Gaussian noise to the original signal, i. e., we set $\bar{q}^i = \exp_{q^i} X_i$ for $i = 1, \ldots, n$, where $X \in \mathcal{T}_{q^i}\mathcal{M}$ has a standard deviation of $\sigma = 0.1$.

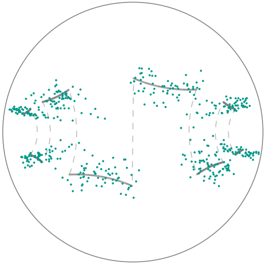


Figure: Artificial signal $q \in (\mathcal{H}^2)^{496}$ in gray, and noisy data $\bar{q} \in (\mathcal{H}^2)^{496}$ in teal.

Numerical Example: Signal Denoising on $\mathcal{H}^2$

Let $\mathcal{H}^2$ be the two-dimensional hyperbolic space and let $n = 496$. By projecting a square wave onto $\mathcal{H}^2$, we manufacture an artificial signal which can be interpreted as a point $q \in (\mathcal{H}^2)^n$. We then generate a noisy signal $\bar{q} \in \mathcal{M}$ by adding Gaussian noise to the original signal, i. e., we set $\bar{q}^i = \exp_{q^i} X_i$ for $i = 1, \ldots, n$, where $X \in \mathcal{T}_{q^i}\mathcal{M}$ has a standard deviation of $\sigma = 0.1$. We want to solve the following total variation problem

$$f_q(p) = \frac{1}{n}\big(g(p, q) + \alpha\,\mathrm{TV}(p)\big)$$

where

$$g(p, q) = \frac{1}{2}\sum_{i=1}^{n} \mathrm{dist}(p^i, q^i)^2 \quad \text{and} \quad \mathrm{TV}(p) = \sum_{i=1}^{n-1} \mathrm{dist}(p^i, p^{i+1}).$$

We also set $\mathrm{diam}(\mathrm{dom}\, f) := \frac{5}{2}\,\mathrm{dist}(q, \bar{q})$.

Numerical Example: Signal Denoising on $\mathcal{H}^2$

Finally, we compare the Riemannian Convex Bundle Method (RCBM) to the Proximal Bundle Algorithm (PBA), the Subgradient Method (SGM), and the Cyclic Proximal Point Algorithm (CPPA).

Numerical Example: Signal Denoising on $\mathcal{H}^2$

Finally, we compare the Riemannian Convex Bundle Method (RCBM) to the Proximal Bundle Algorithm (PBA), the Subgradient Method (SGM), and the Cyclic Proximal Point Algorithm (CPPA).

Algorithm	Iter.	Time (sec.)	Objective	Error
RCBM	3232	94.290	$1.7929 \cdot 10^{-3}$	$3.3101 \cdot 10^{-4}$
PBA	14 879	79.036	$1.8160 \cdot 10^{-3}$	$4.2182 \cdot 10^{-4}$
SGM	15 000	78.742	$1.7918 \cdot 10^{-3}$	$3.3004 \cdot 10^{-4}$
CPPA	15 000	73.065	$1.7928 \cdot 10^{-3}$	$3.3229 \cdot 10^{-4}$

Comparisons between the four algorithms on $(\mathcal{H}^2)^{496}$ for a TV parameter of $\alpha = 0.05$.

Numerical Example: Signal Denoising on $\mathcal{H}^2$

Finally, we compare the Riemannian Convex Bundle Method (RCBM) to the Proximal Bundle Algorithm (PBA), the Subgradient Method (SGM), and the Cyclic Proximal Point Algorithm (CPPA).

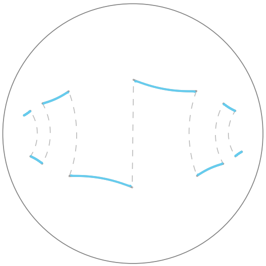


Figure: Denoised reconstruction $q^* \in (\mathcal{H}^2)^{496}$ in cyan.

Numerical Example: Riemanniann Median on $\mathcal{S}^d$

Let $\mathcal{S}^d$ be the $d-$dimensional sphere and let $q^1, \ldots, q^n \in \mathcal{S}^d$ be $n = 1000$ Gaussian random data points sampled in a ball of radius $\frac{\pi}{3}$ around the north pole $\bar{p}$, $B_{\frac{\pi}{3}}(\bar{p})$.

NTNU

Numerical Example: Riemanniann Median on $\mathcal{S}^d$

Let $\mathcal{S}^d$ be the $d-$dimensional sphere and let $q^1, \dots, q^n \in \mathcal{S}^d$ be $n = 1000$ Gaussian random data points sampled in a ball of radius $\frac{\pi}{3}$ around the north pole $\bar{p}$, $B_{\frac{\pi}{3}}(\bar{p})$. Let $f \colon \mathcal{S}^d \to \overline{\mathbb{R}}$ be defined by

$$f(p) = \begin{cases} \frac{1}{n} \sum_{j=1}^{n} \operatorname{dist}(p, q^j) & \text{if } p \in B_{\frac{\pi}{3}}(\bar{p}), \\ +\infty & \text{otherwise.} \end{cases}$$

NTNU

Numerical Example: Riemanniann Median on $\mathcal{S}^d$

Let $\mathcal{S}^d$ be the $d-$dimensional sphere and let $q^1, \ldots, q^n \in \mathcal{S}^d$ be $n = 1000$ Gaussian random data points sampled in a ball of radius $\frac{\pi}{3}$ around the north pole $\bar{p}$, $B_{\frac{\pi}{3}}(\bar{p})$. Let $f\colon \mathcal{S}^d \to \overline{\mathbb{R}}$ be defined by

$$f(p) = \begin{cases} \frac{1}{n}\sum_{j=1}^{n} \operatorname{dist}(p, q^j) & \text{if } p \in B_{\frac{\pi}{3}}(\bar{p}), \\ +\infty & \text{otherwise.} \end{cases}$$

Consider now the dataset

$$\mathcal{D} = \left\{ q^1, \ldots, q^n \,\middle|\, q^j \in \mathcal{S}^d \text{ for all } j = 1, \ldots, n \right\}.$$

Numerical Example: Riemanniann Median on $\mathcal{S}^d$

Let $\mathcal{S}^d$ be the $d-$dimensional sphere and let $q^1, \dots, q^n \in \mathcal{S}^d$ be $n = 1000$ Gaussian random data points sampled in a ball of radius $\frac{\pi}{3}$ around the north pole $\bar{p}$, $B_{\frac{\pi}{3}}(\bar{p})$. Let $f\colon \mathcal{S}^d \to \overline{\mathbb{R}}$ be defined by

$$f(p) = \begin{cases} \frac{1}{n}\sum_{j=1}^{n} \operatorname{dist}(p, q^j) & \text{if } p \in B_{\frac{\pi}{3}}(\bar{p}), \\ +\infty & \text{otherwise.} \end{cases}$$

Consider now the dataset

$$\mathcal{D} = \left\{ q^1, \dots, q^n \,\middle|\, q^j \in \mathcal{S}^d \text{ for all } j = 1, \dots, n \right\}.$$

The Riemannian geometric median p^* of $\mathcal{D}$ is then defined as

$$p^* := \operatorname*{arg\,min}_{p \in \mathcal{S}^d} f(p).$$

The initial point is chosen as one of the two points that realize the maximal distance within $\mathcal{D}$.

Numerical Example: Riemanniann Median on $\mathcal{S}^d$

	RCBM			PBA		
Dimension	Iter.	Time (sec.)	Objective	Iter.	Time (sec.)	Objective
2	16	$4.74 \cdot 10^{-3}$	0.393 47	31	$7.80 \cdot 10^{-3}$	0.393 47
4	77	$2.85 \cdot 10^{-2}$	0.391 46	71	$2.29 \cdot 10^{-2}$	0.391 46
32	20	$1.32 \cdot 10^{-2}$	0.393 48	19	$1.09 \cdot 10^{-2}$	0.393 48
1024	30	$3.06 \cdot 10^{-1}$	0.402 59	42	$4.16 \cdot 10^{-1}$	0.402 59
32 768	50	$1.85 \cdot 10^{1}$	0.391 92	78	$2.64 \cdot 10^{1}$	0.391 92

	SGM		
Dimension	Iter.	Time (sec.)	Objective
2	4570	$2.87 \cdot 10^{-2}$	0.393 47
4	5000	1.33	0.391 46
32	5000	2.16	0.393 48
1024	4646	$4.31 \cdot 10^{1}$	0.402 59
32 768	75	$2.26 \cdot 10^{1}$	0.391 92

Comparisons between the three algorithms on S^d with varying dimension.

Conclusion and Future Work

In summary:

- presented the bundle method for geodesically convex functions on Riemannian manifolds
- touched upon convergence and the problems therein
- showed two numerical examples

To do:

- further investigate convergence

Bibliography

- Alimisis, Foivos et al. (2021). "Momentum improves optimization on Riemannian manifolds". In: *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*. Ed. by Arindam Banerjee and Kenji Fukumizu. Vol. 130. Proceedings of Machine Learning Research. PMLR, pp. 1351–1359. URL: https://proceedings.mlr.press/v130/alimisis21a.html.
- Axen, Seth D. et al. (2021). *Manifolds.jl: An Extensible Julia Framework for Data Analysis on Manifolds*. arXiv: 2106.08777.
- Bergmann, Ronny (2022). "Manopt.jl: Optimization on Manifolds in Julia". In: *Journal of Open Source Software* 7.70, p. 3866. DOI: 10.21105/joss.03866.
- Bonnans, J.-F. et al. (2006). *Numerical optimization: theoretical and practical aspects*. Springer-Verlag.
- Geiger, C. and C. Kanzow (2002). *Theorie und Numerik restringierter Optimierungsaufgaben*. New York: Springer. DOI: 10.1007/978-3-642-56004-0.